



## CIP's Open Data & Data Management Guidelines and Procedures

### 1.1 Scope

The CIP Data Management Guidelines and Procedures aim to provide guidance and support throughout the Data Management Cycle to facilitate the implementation of the CIP Open Data and Data Management Policy. This guidance is provided via recommended protocols and processes to ensure quality of final data and timely upload to the institutional Open Access repository.

### 1.2 Data Management throughout the Project Life Cycle

Adopting the Project Life Cycle to frame data management (Data Life Cycle) allows showing in an orderly manner, how the data workflow is connected to the project. This in turn should facilitate the management of the activities within the “Data Life Cycle” (Figure 1).

#### a) Proposal Development (Resource Mobilization)

Funding agencies increasingly require that project proposals include data management plans. During this early stage, researchers need to take into account important variables when planning how they will generate and manage their data, which can later simplify the process and make it more efficient. Suggested activities for this stage are:

- Familiarize yourself with your donor Open Access Policy, as it may contain some parameters that are different to CIP's.
- Conduct a review of related existing data sets in the institutional archives (memory), to identify whether existing data sets can be of use to your project. Refer to Annex 8 (Data Repositories).
- Determine whether a new dataset will be produced.
- Prepare a Data Management Plan, at this stage you should have a clear understanding of the type of protocol you will use in your trial. Refer to Annex 1 (Research Protocols) and to Annex 2 (Data Management Plan Template).
- Budget the costs related to the Data Management Plan including costs for curation, anonymization, etc. If applicable (remember that for all activities related to data management you can seek the support from RIU). Refer to Annex 3 (Annex 3. Budgeting and Planning Template).

#### b) Contract Management (Grants & Contracts)

These are the suggested steps to consider regarding Data Management when celebrating a new agreement (contract, LOU, etc)

- Understand funding agency requirements for Data Management.
- Ensure clarity of ownership of data according to CIP's Policy, especially with sub-grantees (implementing partners)

- Develop the rights and obligations of all parties with respect of their roles and responsibilities for the management and retention of research data.
- Contractual language should cover the following topics:
  - ✓ Confidentiality and disclosure.
  - ✓ Ownership and licensing of data.

Please seek advice from G&C, by consulting with the Compliance and Intellectual Assets Manager.

### c) Project Start-up

In the project start-up meeting, the PI will share the Data Management Plan. G&C should capture in the Compliance Matrix, the research outputs/information products that will result from the project activities, highlighting whether peer reviewed publications and/or data will be generated. If this is the case, and data sets will be an output of the project, RIU will open a project tracker, and will engage with the PI to offer support and/or guidance.

### d) Project On-going

These are the suggested phases to consider at the project On-going stage. The PI and his/her team together with RIU must ensure best practices are followed in each of these steps.

- Phase 1. Creating data (research design, data management plan, plan consent for sharing, locate existing data, collect data, capture and create metadata). Refer to Annex 4 (Documentation and Metadata).
- Phase 2. Processing data (enter data, digitalize, transcribe, translate, check, validate, clean , anonymize data where necessary, describe data, manage and store data). Refer to Annex 5 ( (Data Quality)).
- Phase 3. Analysis of data (interpret data, derive data, produce research outputs, author publications, prepare data for preservation)
- Phase 4. Preserving data. Researchers must ensure that all research data, regardless of format and progress status with respect to the above phases is stored securely and backed up or copied regularly (daily basis is recommended). The data must be stored on network drives in the institutional archive, which is managed by RIU and IT. Personal computers, laptops and external storage devices such as USB's, CD's and DVD's are not recommended for research data storage. Field books data must be digitalized as soon as the collection of data is completed. Refer to Annex 6 (Data Storage and Archiving). Special attention should be given to sensitive records/data that must be appropriately protected from unauthorized access. Refer to Annex 7 (Privacy & Confidentiality).

### e) Project Close-out

During the project Close-out phase, RIU will have to confirm whether final datasets have been made OA. If a dataset is expected to be made OA after the project activities have been completed, estimated date for OA must be provided. G&C and RIU will follow up, and the project will be formally closed for CIP only when G&C has received confirmation from RIU that datasets have been made OA. According to CIP's Open Data and Data Management Policy, data as a research output must be made available within 12 months of completion. Data associated with a distinct research output (like a peer reviewed article) must be made available within 6 months of publication of the research output.

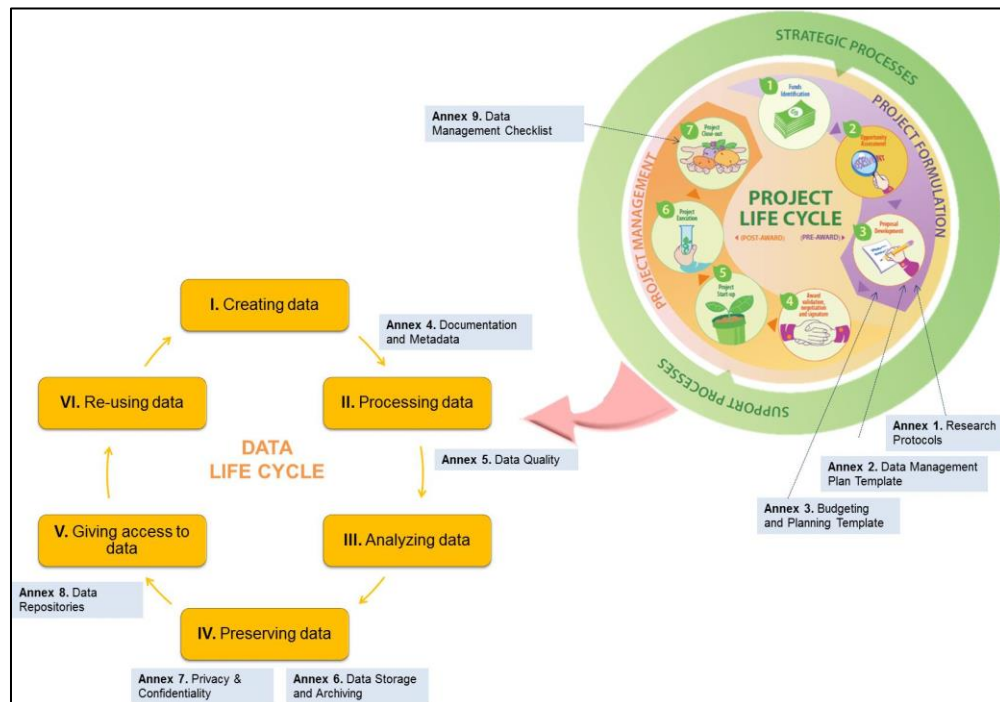
Steps for this stage are as follows:

- Phase 5. Preserving data.
  - ✓ Create metadata and documentation. Refer to Annex 4 (Documentation and Metadata).
  - ✓ Migrate data to best format. Refer to Annex 5 (Data Quality).
  - ✓ Back-up and store data. Refer to Annex 6 (Data Storage and Archiving).

- Phase 6. Giving access to data.
  - ✓ Upload your data set(s) into Dataverse<sup>1</sup>, the institutional repository. Refer to Annex 8. (Data Repositories).

The PI will request assistance from RIU for the process of making data openly accessible. Before final data is archived, additional controls must take place. The PI together with RIU will check for accuracy and appropriateness of formats, quality of metadata, confirm archive, confirm anonymization (sensitive records have been appropriately protected) if defined as a requirement during the start-up meeting. Once Data sets have been finally archived in the OA repository, RIU must give this status to the project data set in the institutional records (tracking system). Please refer to Annex 9. Data Management Checklist.

**Figure 1.** Data Management and the Project Life Cycle.



### 1.3 Data Management and Performance Appraisals

All PI's that manage research projects as well as their direct reports that implement activities related to the Data Life Cycle, must have as one item of their work plans, compliance with CIP Open Data and Data Management Policy and CIP Data Management Guidelines and Procedures. Work plans must include compliance indicators to ensure that quality, best practices, harmonization and standards, punctuality amongst other criteria are promoted within the research community.

### 1.4 Data Management and Staff Entry and Exit procedures

Every new staff member that joins CIP's research community must be familiar with these guidelines and associated policies. Every PI must ensure that the staff working in a project understand the content of these guidelines and are capable of complying with them. Orientation programs of new staff will include proper alignment with CIP Publications Policy, CIP Open Data and Data Management Policy, and CIP Open Data and Data Management Guidelines and Procedures.

<sup>1</sup> <https://data.cipotato.org/>

The supervisor is responsible for requesting and ensuring proper orientation. HR is responsible for including OA as part of the orientation process. Training sessions will also be programmed for existing staff; these will be implemented by RIU and the Library on Open Data and Data Management and Publications respectively.

When a staff member is leaving CIP, HR will include as part of the standard exit procedure a section on Data and other Information Products. All the information products, which resulted from research activities performed by the staff member, constitute CIP's intellectual property, and as such must be cleared by RIU Unit.

### 1.5 Roles and Responsibilities

Clarity in roles and responsibilities is necessary to guarantee sound management of the Data Life Cycle.

- a) **The DDG-RD** is responsible for enforcing these guidelines and for ensuring a continued improvement process.
- b) **HR** is responsible for including Open Access and Open Data as part of the Orientation to new staff, and for including in staff contracts the information regarding staff responsibilities on Open Access and Open Data.
- c) **G&C** is responsible for including the relevant language in sub-grantees (collaborating partners) agreements to guarantee that CGIAR as well as CIP and donor (whichever more stringent) policies and standards cascade down.
- d) **Principal Investigators (PI)** as project leaders have overall accountability for managing research data along the Project Life Cycle and must ensure that a Data Management Plan at the project level is developed and implemented, delegating specific responsibilities as appropriate. The PI is the project designated data focal point and she/he is responsible for coordinating with RIU the required support to deliver data sets of high quality, in accordance with institutional standards and made openly accessible in a timely manner according to the CIP Open Data & Data Management Policy.
- e) **Researchers.** Project researchers are responsible for implementing the Data Life Cycle stages and activities, as defined in these guidelines. Collection of data, maintenance of records, data back-up, data metadata, and archiving amongst other activities must follow best practices and standards described in these guidelines.
- f) **Partners.** Information products produced by CIP and partners are subject to the Policy and Guidelines on all new contracts established since the adoption of the Policy. A contract may have been entered into, which contains restrictions on, for example, sharing the data under a research and/or development project or under a commercialization endeavor. Future agreements should be carefully negotiated to ensure that any such restrictions are limited in duration, territory and/or field of use, if applicable, and fully justifiable by reference to the CGIAR Principles on the Management of Intellectual Assets<sup>2</sup> (i.e., in particular articles 6.2, 6.3 and 6.4) and these guidelines and procedures.

### 1.6 Annexes

Most data often have a much longer lifespan than the research project that creates them. Researchers may continue to work on data after funding has ceased, follow-up projects may analyze or add to the data, or data may be reused and repurposed by other researchers. If data are well managed during the course of a research project, and if they are properly preserved, curated and made accessible for the longer term, they will be able to be reused in future research.

This guideline is based on the UK data archive model for Data Life Cycle management<sup>3</sup>. It consists of six phases: creating, processing, analyzing, preserving, giving access, and re-using data. Phases 1 to 5 are addressed as an integral part of the Project Life Cycle and described in the Annexes. The Guideline has 9 annexes. A description of each annex is detailed on Table 1.

---

<sup>2</sup> See the CGIAR IA Principles on the Management of Intellectual Assets  
<https://library.cgiar.org/bitstream/handle/10947/3755/CGIAR%20IA%20Principles.pdf> for more details.

<sup>3</sup> <http://www.data-archive.ac.uk/create-manage/life-cycle>.

**Table 1.** List of Annexes of CIP’s Open Data & Data Management Guidelines and Procedures

List of annexes	Description
Annex 1. Research Protocols	A research protocol is a written document in which every researcher should have the details exactly how the activity has to be carried out. The annex describes the content and structure of a research protocol and gives examples of published manuals, guidelines and operational procedures.
Annex 2. Data Management Plan Template	This annex contains CIPs standard template for the Data Management Plan (DMP) of a project
Annex 3. Budgeting and Planning Template	This annex contains a budget template to facilitate costing for Open Access and Open Data. It identifies all significant expenses that may be applicable to a project.
Annex 4. Documentation and Metadata	At the basic level metadata is information that enables the user to fully understand a dataset. It is often described as “data about data”. The metadata facilitates discovery of open data sets and therefore reuse.  The annex describes the Core Metadata Schema, the Core Data Dictionary adopted as CIP’s standards and a link to main data dictionaries available at the present for further use.
Annex 5. Data Quality	Quality assurance of data has two dimensions: Ex-ante data quality is defined as quality based on a research protocol. Ex-post quality is defined as quality based on consistency with established internal and external data standards.  The annex describes data quality assurance and guidance for handling ‘special’ types of data.
Annex 6. Data Storage and Archiving	All data generated as a result of research funded by CIP programs needs to be deposited in a suitable repository. The data storage and archiving is basically an area where the project data and documents are stored and the rules that enable a team to use it effectively.  The annex addresses the storage and preservation issues, the use of ISA-tab standards for folder structure and naming files, and finally, the process of maintaining research data from projects so that it can be found and used in the future.
Annex 7. Privacy & Confidentiality	Researchers are responsible for the ethical treatment of data. Research data which includes confidential or private information must be managed in accordance with any contractual or funding agreements. The annex describes how Privacy and Confidentiality; the Prior Informed Consent Agreement need to be addressed. The annex outlines how datasets becomes anonymous before data archiving.
Annex 8. Data Repositories	CIP is using a locally-hosted Dataverse installation. Dataverse is a commonly-used open source data repository platform that facilitates the ability to publish, share, reference, extract and analyzes research data. The tool helps to make research data openly accessible.
Annex 9. Data Management Checklist	The Data Management Checklist enumerates all items that need to be completed at project Close-out.

## **1.7 References**

CGIAR (2015). Open Access & Open Data Support Pack. <https://sites.google.com/a/cgxchange.org/oad-support-pack/home>. Accessed on April 27, 2015.

CGIAR (2012). GIAR Principles on the Management of Intellectual Assets (“CGIAR IA Principles”). <https://library.cgiar.org/bitstream/handle/10947/3755/CGIAR%20IA%20Principles.pdf/> Accessed on April 27, 2015.